

Power analysis of escapement trends for the Coho salmon protocol

Report by:

**Leigh Ann Starceвич
University of Idaho, Moscow, Idaho**

**Task Agreement: J8W07060004
Cooperative Agreement: H8W07060001**

**Submitted to the San Francisco Bay Area Network, Golden Gate National
Recreation Area, Sausalito, CA 94965.**

August 2008

San Francisco Area Network (SFAN) monitors Coho salmon populations within parks located in the coastal watersheds of Marin County. Long-term monitoring projects are used to detect changes over time in attributes of the population. Power analyses of the tests of trend are useful in assessing what degree of change may be detected or the length of time needed to detect trend given the current methodology. The power to detect trends of escapement metrics is examined in this report.

Survey Design

Escapement surveys are usually conducted between November and January and depend on environmental conditions conducive to spawning. Monitoring is conducted in 2 to 4 km reaches within Olema, Pine Gulch, Redwood, and Cheda Creeks. Reaches within these watersheds are not randomized and effort may vary across years. Given that the reaches are subjectively chosen, the scope of inference for estimated trends is restricted to these four watersheds only.

SFAN personnel census creek main stems and some selected tributaries. Watersheds are visited once a week or less if water is high or turbid. Teams of two to four observers walk upstream and count live fish, carcasses, and redds. Live fish are measured visually and sex and species is recorded. Carcasses are measured and sexed to prevent double counting. Redds are flagged and length and width measurements are recorded. From these data, estimates of Area Under the Curve (AUC), Peak Live Cumulative Dead (PLD), and redd counts and density may be obtained.

As with many biological resources, detection error can affect an observer's ability to census a population. If detection probabilities do not differ dramatically over time, then the ability to detect trend over time for these four watersheds may not be compromised by detection error.

Pilot Data

The pilot data used in this trend example are redd counts from escapement surveys conducted from the 1997-98 to the 2006-2007 spawner season in Olema, Pine Gulch, and Redwood Creeks. Reaches were visited between once and 10 times within a season. Since redds are flagged and are unlikely to be double-counted, observed redds were totaled to watershed level and density estimates were calculated using the total stream length surveyed annually within a watershed. Redds recorded as "definite" redds were used for Coho only.

The Olema Creek watershed is extensive and cannot be surveyed in a single day. PLD estimates are time-dependent, so these estimates are calculated for tributaries surveyed during the same day. Olema Creek Mainstem, John West Fork, and a group of four tributaries (Boundary Gulch, Giacomini Creek, Horse Camp Creek, and Quarry Gulch) each have separate PLD measurements. A weighted average of Olema Creek mainstem, John West Fork, and the 4 smaller tributaries was computed to summarize PLD at the watershed level and to avoid double counting errors. The weighted mean is a conservative summary of PLD because it is always less than the maximum PLD measurement.

Pilot data sample sizes of years used in the trend analysis are given in Table 1. All four watersheds were visited annually since 2000.

Table 1: Sample sizes per watershed and year

Watershed	Number of years for PLD measurements
Cheda	9
Olema	10
Pine Gulch	7
Redwood Creek	10

Figures 1 and 2 display the PLD and redds per km metrics, respectively, calculated in each watershed by survey year. Notice the strong cohort effect, with cohort 1 having higher values for both measures, cohort 3 with lower values, and cohort 2 falling in between.

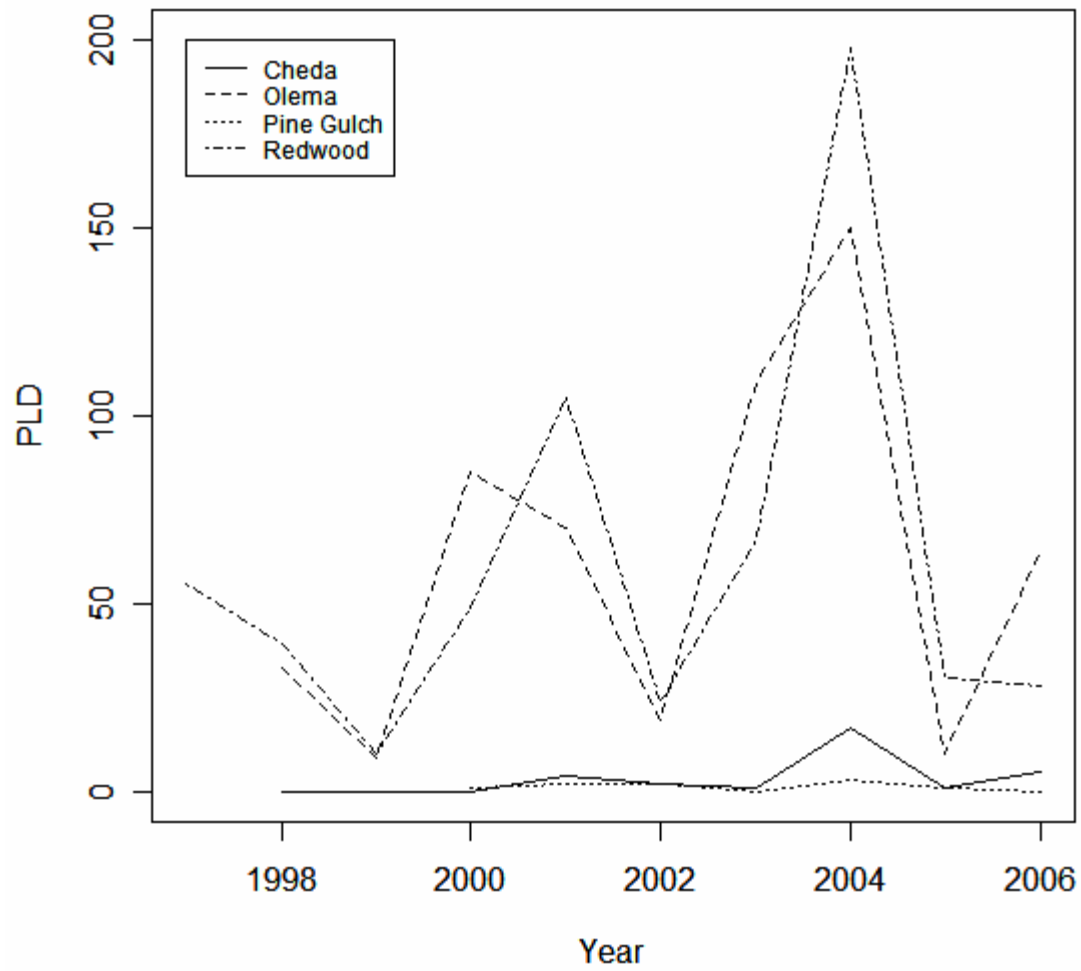


Figure 1: PLD by Year and Watershed

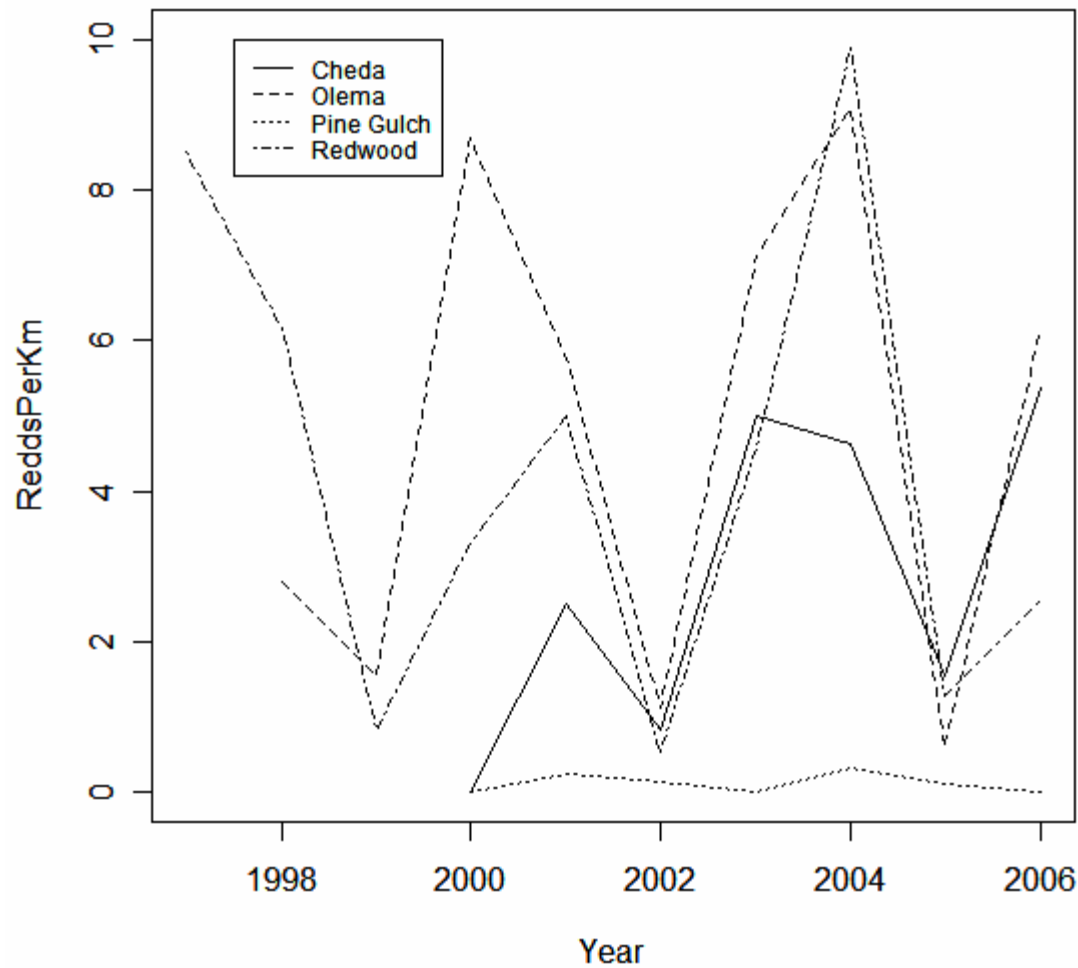
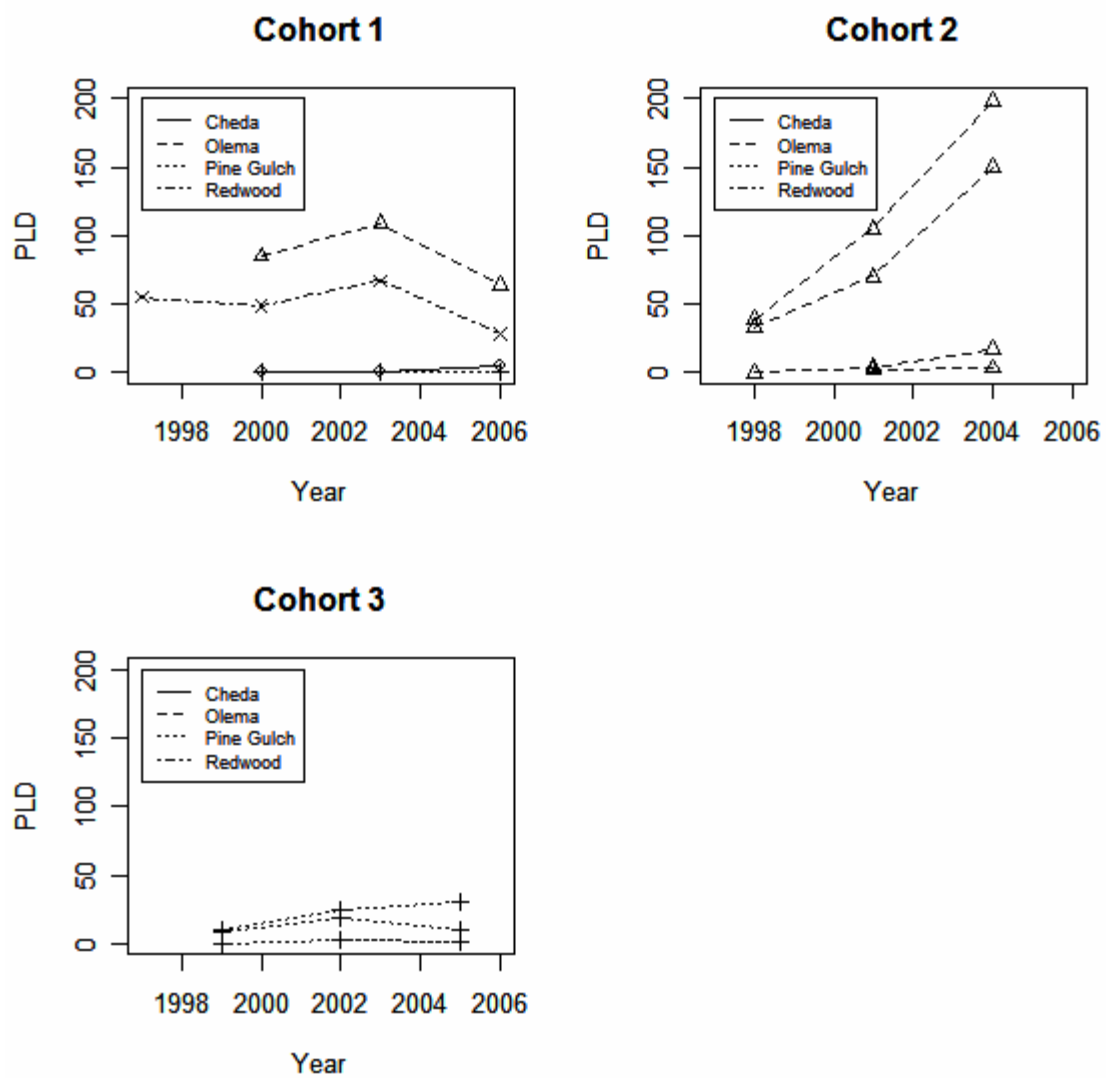
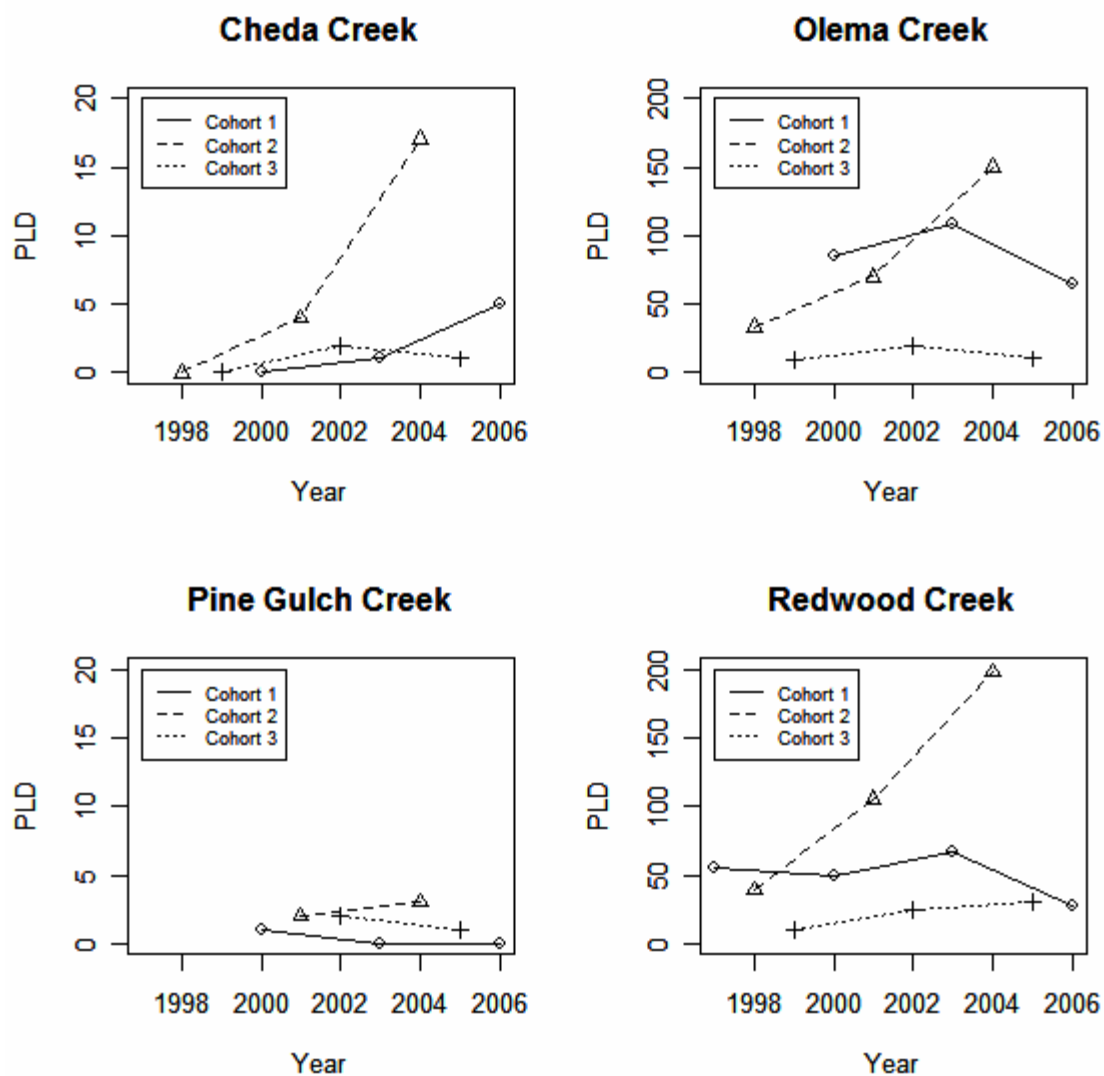


Figure 2: Redds per Km by Year and Watershed

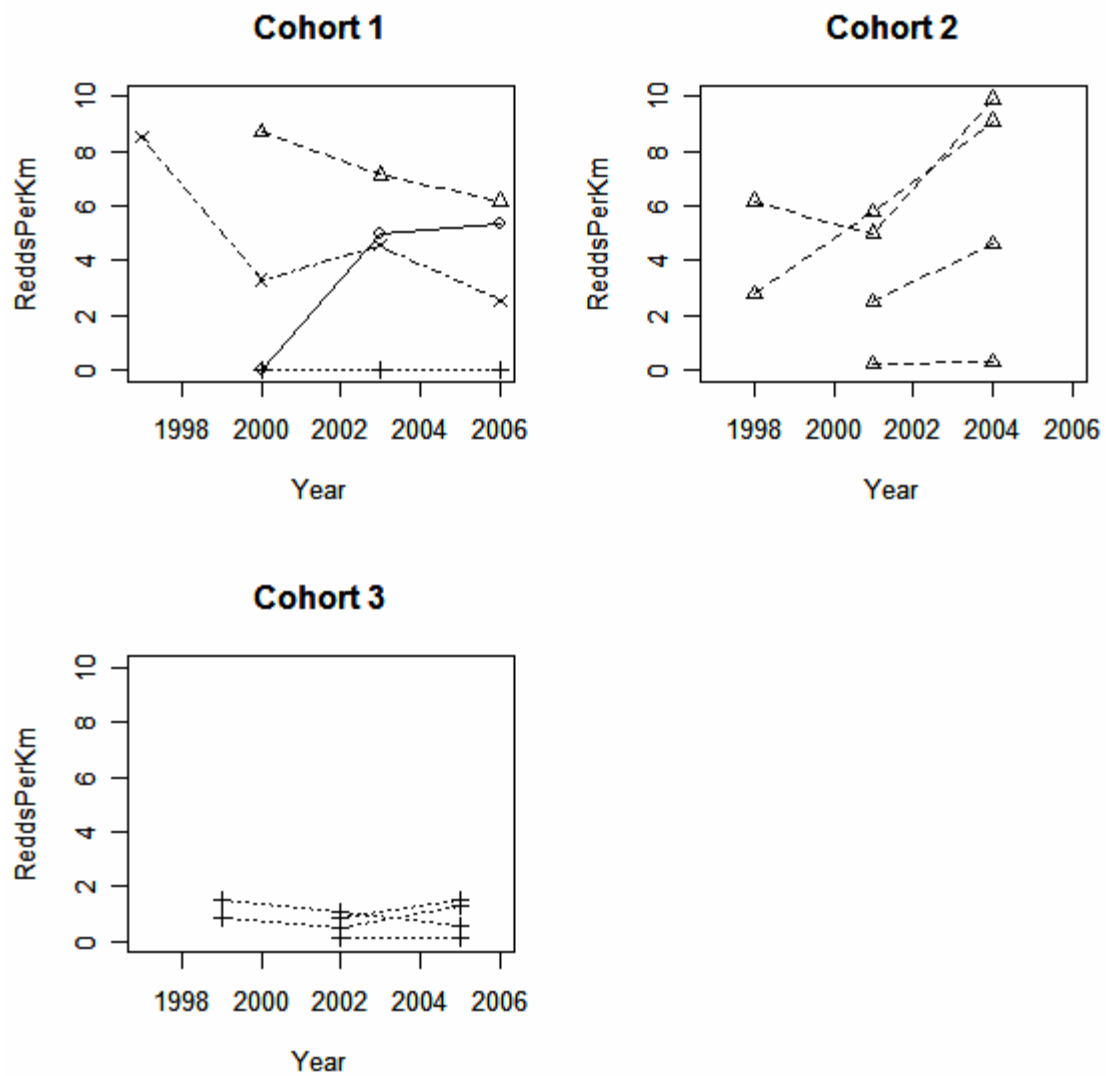
True replication for this system is found within a cohort and watershed. Despite the fact that all watersheds are visited annually, true replication is achieved every three years due to the nature of the Coho return cycle. The following figures provide profile plots of both outcomes for watersheds within cohorts and for cohorts within watersheds. An interaction between watershed and cohort is more evident in the plots of cohorts within watersheds.



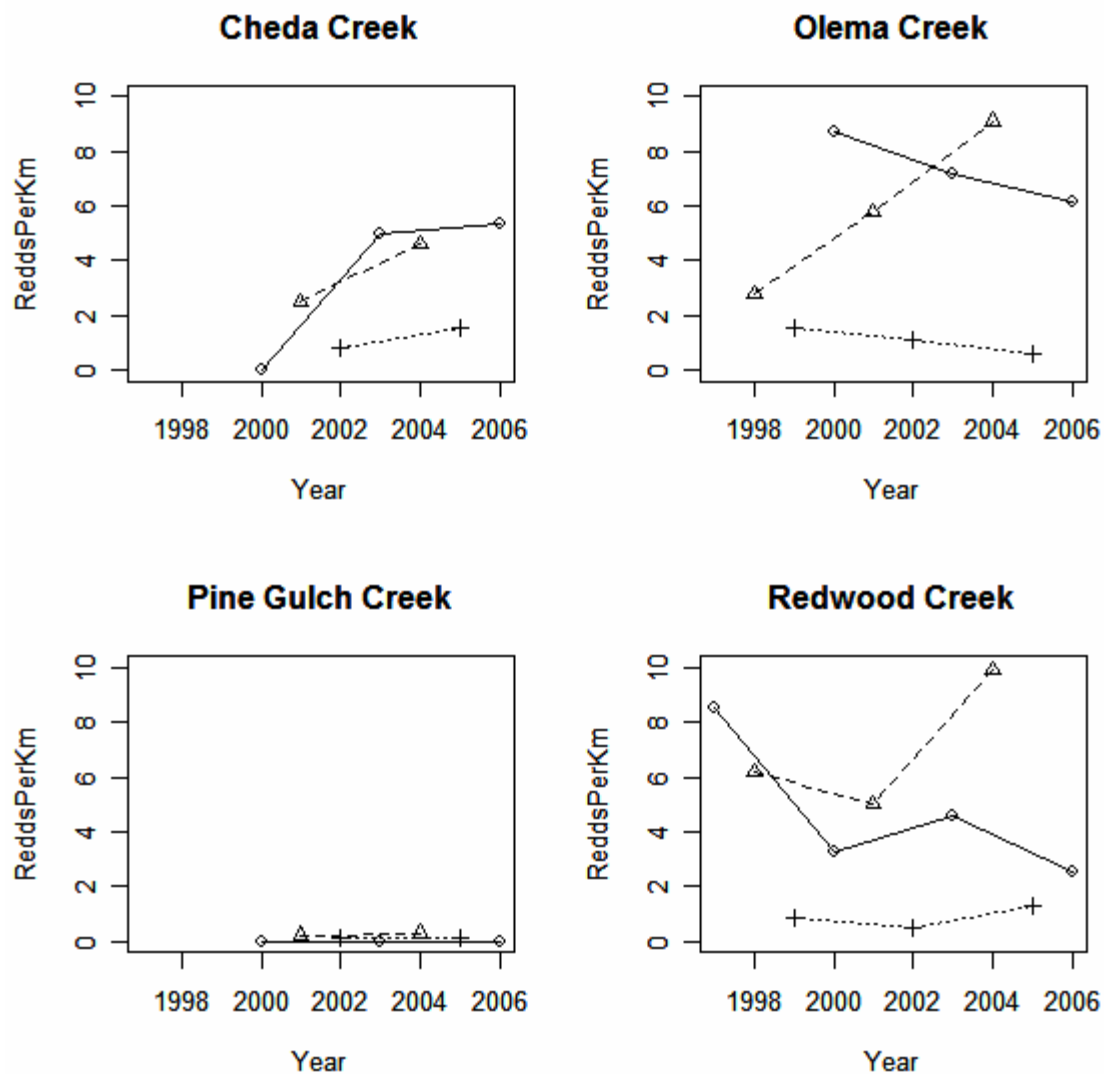
Figures 3 – 5: PLD measures by watershed for cohorts 1 - 3



Figures 6 – 9: PLD measures by cohort for all four watersheds



Figures 10 – 12: Redds per km measures by watershed for cohorts 1 - 3



Figures 13 – 16: Redds per km measures by cohort for all four watersheds

Trend model selection

Piepho and Ogutu (2002) parameterize year as both a fixed effect for testing linear trend and as a random effect to account for annual variation in the outcome. The fixed effect of year, referred to as *WYear* in their paper, is defined as an integer starting at 0 and increasing each year by one. Because SFAN personnel want to interpret three-year multiplicative changes in outcomes, the fixed effect for year is parameterized as 0, 0, 0, 3, 3, 3, 6, 6, 6, ... to account for the Coho return schedule. This fixed effect for year is referred to here as *CCYear*, for creek-cohort year.

Initial trend analysis with the pilot data suggested that variance components were very high. Several models were examined with the goal of finding a model appropriate for the Coho replication structure that could also explain the extensive variation in the system. A basic trend model as in Piepho and Ogutu (2002) was examined, as well as a model

with a fixed interaction effect between watershed and cohort and a model treating the cohort effect as random.

Piepho and Ogutu (2002) use REML estimation for model parameters. Gurka (2006) suggests that the Bayes Information Criterion (BIC) is more appropriate for model selection of mean structures under REML. For the following models, Table 2 provides BIC values.

Model 1: No interaction effect for watershed-by-cohort

Model 2: Interaction effect for watershed-by-cohort

Model 3: Random intercept for cohort

Table 2: BIC values for three models and both outcomes

Model	PLD	Redds per km
1	70.5	51.1
2	44.1	34.3
3	74.7	55.2

For both outcomes, the BIC values suggest that the model including the fixed interaction effect between watershed and cohort is superior in explaining variation in escapement outcomes.

Trend Model

The mixed linear model of VanLeeuwen, et al. (1996) and Piepho and Ogutu (2002) is modified to reflect the cohort structure of Coho salmon. Assuming one measure per watershed per year, the modified mixed linear model is given by:

$$y_{ijk} = w_j\beta + \alpha_i + b_j + w_j t_i + \gamma_k + \phi_{ik} + w_j d_k + e_{ijk},$$

where $i=1,\dots,m_a$ indexes watershed;

$j=1,\dots,m_b$ indexes year;

$k=1,2,3$ indexes cohort;

m_a = the number of sites in the sample;

m_b = the number of years in the sample;

w_j = constant representing the j^{th} cohort-by-watershed year;

β = fixed slope of the linear time trend;

α_i = fixed intercept of i^{th} watershed;

b_j = random intercept of the j^{th} year;

t_i = random slope of i^{th} watershed, iid as $N(0, \sigma_t^2)$;

γ_k = fixed intercept of k^{th} cohort;

ϕ_{ik} = fixed interaction effect between the i^{th} watershed and the k^{th} cohort;

d_k = random slope of k^{th} cohort, iid as $N(0, \sigma_d^2)$; and

e_{ijk} = unexplained error, iid as $N(0, \sigma_e^2)$.

The SAS code to obtain the variance components is as follows:

```
proc mixed data=SFAN.Data1Log METHOD=REML;
CLASSES Watershed Cohort Year;
MODEL LogRedd = Watershed Cohort Watershed*Cohort
CCYear/DDFM=SATTERTH residuals s noint;
RANDOM INT/SUB=Year;
RANDOM CCYear/SUB=Watershed s;
RANDOM CCYear/SUB=Cohort s;
run;
```

This linear model forms the basis of the power analysis for the test of the fixed effect for CCYear. Tests were performed at the $\alpha=0.10$ level.

Results

Simulations were conducted in R to generate random effects and assess the power to

detect actual trends over time. The two-sided hypothesis test of no trend was calculated for three-year multiplicative **decreases** of 2.5%, 10%, 25%, 50%, and 75%. The results of the power for tests of trend for PLD and redds per km are given in Figures 17 and 18, respectively.

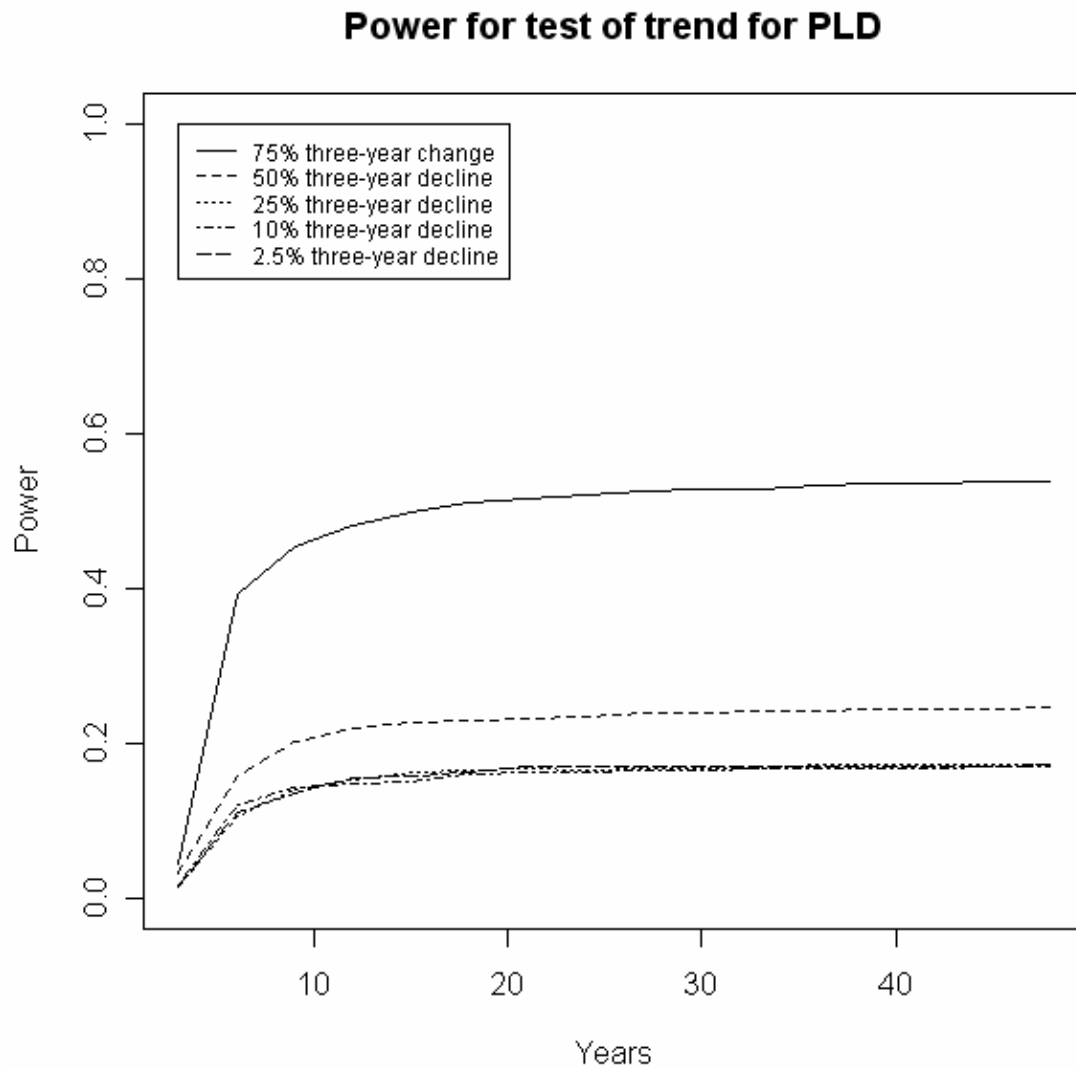


Figure 15

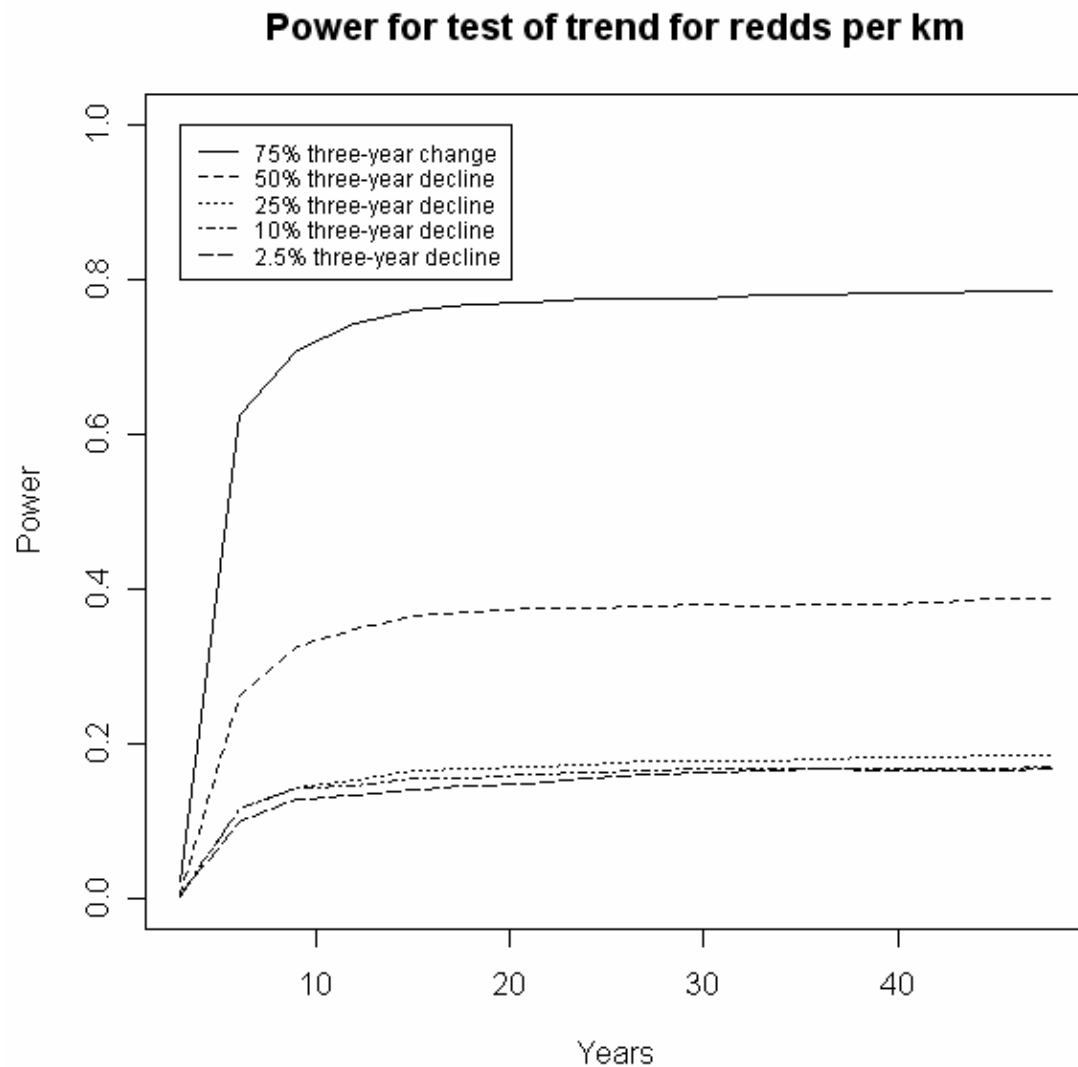


Figure 16

The results of the power analysis indicate that both outcomes are highly variable and require drastic population declines before obtaining the power to detect change with any reliability. Additional environmental covariates may explain some of the variation in cohorts or watersheds or the random unexplained variation.

Figure 4 indicates an increasing trend for cohort 2. The estimates from cohort 2 were used to estimate variance components and conduct a power analysis to determine if the power to detect linear trends in escapement parameters would be greater without the additional sources of variation among cohorts and cohort-watershed interactions. The results of the cohort 2 power analyses for PLD and redds per km are provided in Figures 17 and 18, respectively. The power to detect linear trend in escapement parameters is improved by examining a single cohort, especially for redds per km.

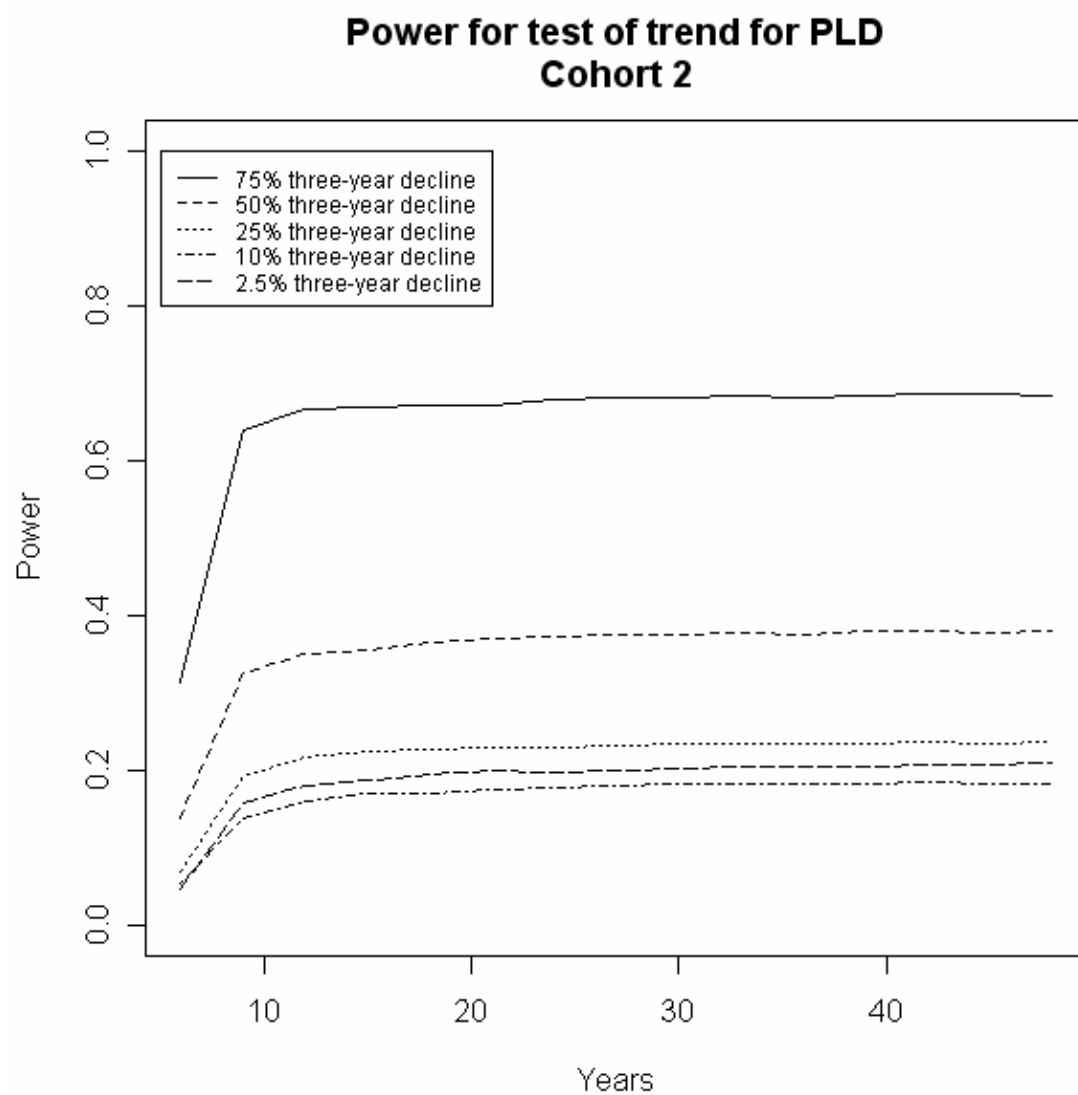


Figure 17

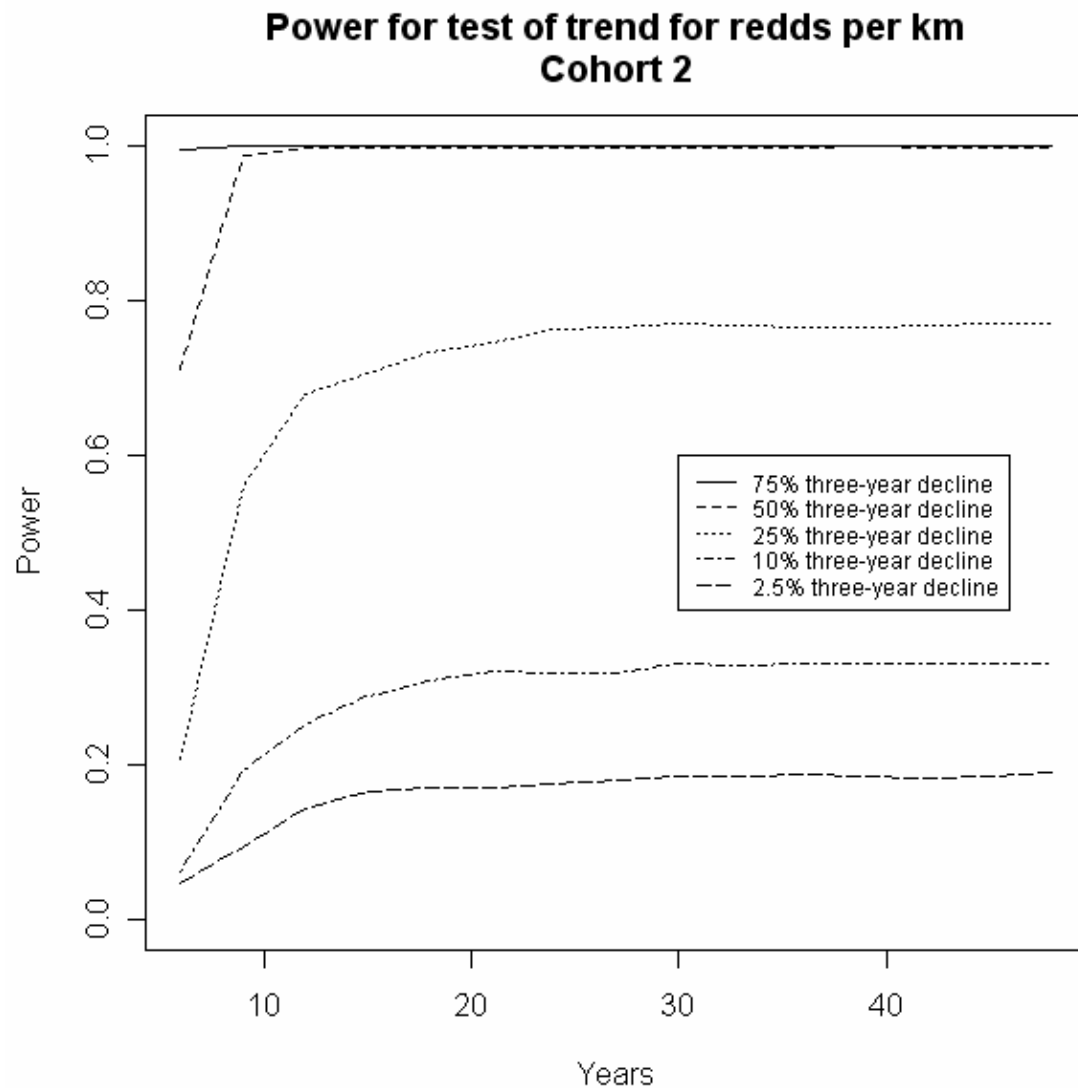


Figure 18

References

- Gurka, M.J. (2006). Selecting the best linear mixed model under REML. *The American Statistician* 60(1): 19-26.
- Piepho, H.P. and J.O. Ogutu (2002). A simple mixed model for trend analysis in wildlife populations. *Journal of Agricultural, Biological, and Environmental Statistics* 7(3): 350-360.
- VanLeeuwen, D.M., L.W. Murray, and N.S. Urquhart (1996). A mixed model with both fixed and random trend components across time. *Journal of Agricultural, Biological, and Environmental Statistics* 1(4): 435-453.